Iniziato	Thursday, 13 January 2022, 15:12
Stato	Completato
Terminato	Thursday, 13 January 2022, 15:41
Tempo impiegato	29 min. 29 secondi
Punteggio	13,00/15,00
Valutazione	**26,00** su un massimo di 30,00 (**87**%)

Domanda **1**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following is not a property of a *metric* distance function

Scegli un'alternativa:

- a. Boundedness ✔
- b. Symmetry
- c. Triangle inequality
- d. Positive definiteness

Risposta corretta.

La risposta corretta è: Boundedness

Domanda **2**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Given the two binary vectors below, which is their similarity according to the Simple Matching Coefficient?

a b c d e f g h i j
1 0 0 0 1 0 1 1 0 1
1 0 1 1 1 0 1 0 1 0

Scegli un'alternativa:

- a. 0.2
- b. 0.3
- c. 0.1
- d. 0.5 ✔

Risposta corretta.

La risposta corretta è: 0.5

Domanda **3**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following statements is *true*?

Scegli una o più alternative:

- a. The noise always generate outliers
- b. The data which are similar to the majority are never noise
- c. The noise can generate outliers ✔
- d. Outliers can be due to noise ✔

Your answer is correct.

Le risposte corrette sono: Outliers can be due to noise, The noise can generate outliers

Domanda **4**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Given the definitions below:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

which of the formulas below computes the accuracy of a binary classifier?

Scegli un'alternativa:

a. TN / (TN + FP)

b. (TP + TN) / (TP + FP + TN + FN) ✔

c. TP / (TP + FP)

d. TP / (TP + FN)

Risposta corretta.

La risposta corretta è: (TP + TN) / (TP + FP + TN + FN)

Domanda **5**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

What is the *cross validation*

Scegli un'alternativa:

- a. A technique to obtain a good estimation of the performance of a classifier with the training set
- b. A technique to improve the quality of a classifier ✖
- c. A technique to obtain a good estimation of the performance of a classifier when it will be used with data different from the training set
- d. A technique to improve the speed of a classifier

Risposta errata.

La risposta corretta è: A technique to obtain a good estimation of the performance of a classifier when it will be used with data different from the training set

Domanda **6**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

A Decision Tree is...

Scegli un'alternativa:

- a. A tree-structured plan of tests on single attributes to forecast the cluster
- b. A tree-structured plan of tests on multiple attributes to forecast the target
- c. A tree-structured plan of tests on single attributes to forecast the target
- d. A tree-structured plan of tests on single attributes to obtain the maximum purity of a node ✖

Risposta errata.

La risposta corretta è: A tree-structured plan of tests on single attributes to forecast the target

Domanda **7**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

When training a neural network, what is the *learning rate*?

- a. The speed of convergence to a stable solution during the learning process
- b. The ratio between the size of the hidden layer and the input layer of the network
- c. The slope of the activation function in a specific node
- d. A multiplying factor of the correction to be applied to the connection weights ✔

Your answer is correct.

La risposta corretta è: A multiplying factor of the correction to be applied to the connection weights

Domanda **8**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What measure is maximised by the Expectation Masimisation algirithm for clustering?

Scegli un'alternativa:

- a. The likelihood of an example
- b. The likelihood of an attribute, given the class label
- c. The support of a class
- d. The *likelihood* the distributions, defined by the parameters found, given the data available ✔

Your answer is correct.

La risposta corretta è: The *likelihood* the distributions, defined by the parameters found, given the data available

Domanda **9**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What does K-means try to minimise?

Scegli un'alternativa:

- a. The *separation*, that is the sum of the squared distances of each point with respect to its centroid
- b. The *separation*, that is the sum of the squared distances of each cluster centroid with respect tho the global centroid of the dataset
- c. The *distortion*, that is the sum of the squared distances of each point with respect to its centroid ✔
- d. The *distortion*, that is the sum of the squared distances of each point with respect to the points of the other clusters

Risposta corretta.

La risposta corretta è: The *distortion*, that is the sum of the squared distances of each point with respect to its centroid

Domanda **10**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following characteristic of data can reduce the effectiveness of DBSCAN?

Scegli un'alternativa:

- a. Presence of clusters with different densities ✔
- b. All the variables are the same range of values
- c. Presence of outliers
- d. Clusters have concavities

Your answer is correct.

La risposta corretta è: Presence of clusters with different densities

Domanda **11**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Match the rule evaluation formulas with their names

$$\frac{1 - sup(C)}{1 - conf(A \Rightarrow C)}$$ Conviction ✔

$$sup(A \cup C) - sup(A)sup(C)$$ Leverage ✔

$$\frac{conf(A \Rightarrow C)}{sup(C)}$$ Lift ✔

$$\frac{sup(A \Rightarrow C)}{sup(A)}$$ Confidence ✔

Your answer is correct.

La risposta corretta è: $\frac{1 - sup(C)}{1 - conf(A \Rightarrow C)}$ → Conviction,

$sup(A \cup C) - sup(A)sup(C)$ → Leverage,

$$\frac{conf(A \Rightarrow C)}{sup(C)}$$ → Lift, $$\frac{sup(A \Rightarrow C)}{sup(A)}$$ → Confidence

Domanda **12**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Consider the transactional dataset below

ID	Items
1	A,B,C
2	A,B,D
3	B,D,E
4	C,D
5	A,C,D,E

Which is the *confidence* of the rule A,C ⇒ B?

Scegli un'alternativa:

- a. 100%
- b. 20%
- c. 40%
- d. 50%

1 / 2

Risposta corretta.

La risposta corretta è: 50%

Domanda **13**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

In a dataset with D attributes, how many subsets of attributes should be considered for feature selection according to an exhaustive search?

Scegli un'alternativa:

a. O(D!)

b. O(D)

c. $O(D^2)$

d. $O(2^D)$ ✔

Risposta corretta.

La risposta corretta è: $O(2^D)$

Domanda **14**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What is the **coefficient of determination R^2**?

a. Measure the amount of error in a linear regression model

b. An index of goodness for a classification model

c. Provide an index of goodness for a linear regression model ✔

d. Measure the amount of error in a regression model

Your answer is correct.

La risposta corretta è: Provide an index of goodness for a linear regression model

Domanda **15**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is different from the others?

Scegli un'alternativa:

- a. Decision Tree ✔ This is not a clustering method
- b. Dbscan
- c. Expectation Maximisation
- d. K-means

Risposta corretta.

La risposta corretta è: Decision Tree

Vai a...

Machine Learning - Python Lab ►